

Assessing team performance – markers and methods

Benjamin Marriage¹, John Kinnear²

¹ Norfolk and Norwich University Hospital, Colney Ln, Norwich NR4 7UY, UK

² Postgraduate Medical Institute, Anglia Ruskin University, Bishop Hall Lane, Chelmsford, UK

Corresponding Author:

John Kinnear

Postgraduate Medical Institute, Anglia Ruskin University, Bishop Hall Lane, Chelmsford, CM1 1SQ, UK

Email: john.kinnear@southend.nhs.uk

Word count: 2814

Abstract

Optimal team performance in healthcare is vital to avoid error. Assessing teams improves their performance by providing accurate feedback to team members. To construct an assessment tool that is both valid and reliable, it is necessary to understand what should be measured, when it should be measured and the context of the team performance. Current assessment methods include observation of behavioural markers, self-assessment by team members, event-based coding, and narrative field notes. Future assessment tools should incorporate best practice elements such as having a theoretical or empirical basis for teamwork competencies, measuring process rather than outcomes, analysing rather than describing performance, capturing non-observable cognitions, distinguishing individual from collective team behaviours, using trained raters, and linking assessment to learning objectives.

Key Words

Teams, Teamwork, Human Factors, Assessment

1. Introduction

Traditionally we have thought the quality of healthcare to be dependent on an individual doctor's expertise. However, modern healthcare more closely resembles complex systems, which for successful functioning rely on teams performing several interdependent actions competently at the same time. Single-step task performance with a linear input-output relationship may be accurately attributed to an individual's action, but anything more complex requires several people to perform in concert. In the complex interactions of real life, competence becomes a shared attribute and is dependent on competence of the collective [1]. In other words, 'patient care is a team sport' [2].

The term 'team' is a broad construct that may include anything from strategic teams, management teams, through to clinical healthcare teams. There are many definitions of what constitutes a team, but the principal features are that it is made up of two or more people, there is dynamic interaction between the individuals, there is interdependency and a common goal, there are specific member roles or functions, and it has a limited life span [3]. A team is distinguished from a group by its role differentiation, but more importantly, by its distribution of cognitive load [4]. This article will specifically address dynamic teams dealing with high-stakes situations, which typically possess the characteristics of having defined tasks, but with an unstable membership.

Evidence for the importance of team performance impacting on patient outcome has grown enormously after the Institute of Medicine report *To Err is Human* [5], which emphasised the point that frequently patient safety failures and error resulted from 'systemic' failures and not individual incompetence. There is now little debate that poor team functioning is at the root of most patient harm. This then begs the question of what is optimal team performance so that we are able to teach and propagate it. Which requires it to be assessed.

When making a judgement on the quality of a team's performance, most observers would agree with Justice Stewart's observation (in a different context), 'I know it when I see it' [6], but this superficial evaluation would provide an inadequate basis for improvement where performance is deemed sub-optimal. It also provides no structure on which to build an empirical or theoretic understanding of team functioning. To develop an adequate evidence base we also need to measure it. This article will review the current methods and markers of team performance assessment, but will not only address the question of how to assess it, but also when to assess it (context) and for what reason (purpose). It is not an exhaustive review of every team assessment method devised, but discusses the current views of what constitutes best practice.

2. Why should we assess team performance?

There are three intersecting reasons for wanting to know more about team performance – the emerging understanding of a team's contribution to the error paradigm, the changing nature of healthcare delivery away from an individual-expert dominated model towards a collective expertise one, and a greater appreciation of the healthcare environment becoming a complex milieu. The assessment of team performance is essential for an accurate understanding of how teams work so that processes can be improved for better outcomes.

Early models of error causation arose from industrial prototypes which posited that error arose from linear cause-and-effect pathways [7]. This conceptual model persisted in healthcare as the influential 'Swiss cheese' model propagated by Reason [8]. However, more recent understanding has changed to a view of patient safety and error being linked to 'systems' failure with an integral part of the system being the healthcare team.

Other high reliability industries have been earlier adopters of the concept of 'team training', the industry studied most intensively by anaesthetists being aviation. Human factors teaching began to be developed in the late 1970s, the result of which was Cockpit Resource Management (CRM) in 1979. The first CRM course was run by United Airlines in 1981 and evolved to the second generation CRM (now Crew Resource Management) in the late 80's, which was more team orientated with greater emphasis on authentic aviation operations. Third generation CRM moved to training in more authentic environments, combining human factor training on a functional flight deck, and progressed further to a 4th generation CRM in the 90's, with full mission simulation for specific airlines in which the teams operated.

In parallel with this increasing focus on team performance has been an evolving perspective on the environment in which the teams operated. There has been a growing use of complexity theory as a conceptual framework to explain modern healthcare [9-12], sometimes overwrought and misguided [13]. However, in terms of how teams operate it is a useful framework. It is worth further defining a continuum that operates in a healthcare setting, with 'complicated' and 'complex' at either end. The former has pathways that can be mapped and planned, with predictable outcomes, and the latter has pathways that are interdependent, unpredictable and whose relationships change every time they interact, giving rise to new unpredictable interactions. In healthcare a cardiac arrest team performing resuscitation may be characterised as operating as part of a complicated 'system', whereas dealing with a patient with sepsis-induced multi-organ failure would be complex. The process for cardiac arrest resuscitation is highly protocolised with a limited number of interventions that can be applied (A-B-C), with little latitude for deviation, after which there is a dichotomous outcome (dead or alive). Comparison of a resuscitation team with a NASCAR pit crew [14] is therefore valid because each element of both processes can be finely choreographed to achieve a predictably effective performance. However, the same cannot be said for the acutely ill septic patient who needs a customised team performance each time, which requires flexibility, adaptation and innovation.

3. What should be assessed?

There is not uniform agreement about the terminology of teamwork competencies in the literature [15-17], for example around the terms human factors and non-technical skills, which will be used interchangeably in this article to reflect their common usage in the literature reviewed. However, it is important to distinguish the level at which teamwork is being examined to understand what is being assessed. Broadly, a team may be assessed for the effectiveness of the goals that are achieved (*outcomes*), or for the enacted behaviours (*processes*) it uses to achieve those goals. The entire enactment is usually termed the *performance* [18].

Global performance is the result of *taskwork* and *teamwork*, where the former describes the actions (knowledge, skills and attitudes) of individual team members and the latter the interactions between team members [19]. It has been argued that effectiveness is more the result of *teamwork* since a team of experts is not an expert team, demonstrated in many studies looking at performance of dynamic teams such as flight-deck crews on aircraft carriers [20]. Weick demonstrated that these crews worked as coherent units guided by a shared understanding of their goals, rather than individually competent members. The concept of collective competence was proposed by Boreham [21-23], and fits with an emerging discourse which describes the competence as something that is dynamic, context-sensitive, distributed, interdependent, and evolving, rather than a stable 'state' that can be possessed [24]. This conceptualisation is variously related to the allied concepts of interactive consciousness, group consciousness and collective intelligence [25,26]. However, assessment at the level of single team members does provide an opportunity for targeted feedback for improved effectiveness at the level of individuals.

4. When we should assess

The characteristics of the assessment method will be determined by its primary purpose, be it for measurement only (research), learning (formative) or high stakes decision-making (summative). There may be cross-over between these functions, but tensions may arise for both the assessor and the assessed if goals are divergent. For example, an overall rating of teamwork performance across an observed performance episode may be useful for determining whether a focussed learning intervention has been effective, but will provide little substrate for specific feedback to members for improvement. Effective learning requires the capture and feedback of specific knowledge, skills and behaviours that contribute to good teamwork.

Team performance can be assessed *in vivo* (workplace) or *in vitro* (simulated environment). The former is an uncontrolled environment that may be more amenable to retrospective analysis of a recorded performance, or several raters may be required to distribute the high cognitive load. A simulated environment creates opportunities to direct performance so that particular behaviours can be tested rather than relying on opportunistic learning. The majority of assessment tools are developed for specific clinical situations such as surgery or anaesthetics, or other acute healthcare settings [27]. When using these tools outside their validated contexts it is important to customise them so that they map accurately to the attributes being assessed [28].

5. What methods of assessment are currently used

There is an ever-growing literature on particular assessment methods for team performance and rather than offer an exhaustive list, examples of methods and their characteristics will be presented. As discussed above, choice of a particular method of assessment depends upon what it is to be used for. Table 1 summarises some of the major methods that are described, and each will be briefly discussed in 5 broad categories: behavioural markers (observational rating scales); team self-assessment methods; events-based assessment (event coding); field notes (qualitative assessment); and combined methods [29].

[TABLE 1]

Behavioural markers are observable actions that indicate aspects of a team performance, and usually include a rating scale that makes inferences about the effectiveness of those observed aspects. Most observational methods use behaviourally anchored rating scales [30], or modified Likert-type scales [31]. The attractiveness of assigning scores to easily observed behaviours makes behavioural markers the most popular assessment method. Another potential advantage is that the ratings can provide an effective means of specific feedback on specific teamwork competencies [32]. However, several issues should be considered. The main potential problems with making judgements about observed behaviours are those of validity (measuring what the tool purports to measure) and reliability (getting the same results each time). Reliability is necessary, but not sufficient, for validity, and it seems intuitive that trained raters and increasing experience are essential to achieve both [33]. This has obvious resource implications since assessors need to be recruited, trained and available. Another issue is that different teamwork behaviour may occur at any point or simultaneously during an assessed episode, which demands the question of which behaviours to score and which to not score when there are over 79 different behaviour domains described [27]. Clearly, this may make cognitive overload of raters problematic.

Events based assessment (EBA) is a method developed for simulation-based team assessment to deal with the complexity of authentic clinical interactions [34,35]. Planned critical events are staged to highlight specific teamwork competencies so that raters know which behaviours should occur at which time point. However, each assessment is particular to the scenario designed, making a generic assessment tool difficult [28]. Also, it continues to rely on observation alone, and is therefore unable to capture the tacit cognitive components of teamwork, such as situational awareness and implicit communication. EBA has been used *in vivo*, but is a much more time-consuming process due to the need to produce a predefined set of behaviours to observe against. Whilst these can be tracked using a simple tally, more often these predefined behaviours are transcribed into a coding manual. Trained observers then use a computer based system which can allow a time stamp to be assigned to each event code as it occurs giving more detail on timing and frequency of events [29]. The fact that EBA is both more technically challenging to arrange and more time-consuming may explain why it is less used than pure observation of behavioural markers.

Self-assessment potentially overcomes the shortcoming of not being able to observe tacit behaviours and cognitions, and allows team members to assess the unobservable elements of teamwork. Self-assessment can be done by providing the team with generic questions regarding performance, or may be combined with another method (usually a behavioural marker-based system). The self-assessment may use the same scoring system as the observational tool to allow comparison (and highlight performance gaps) and further facilitate discussion during feedback.

Field notes usually consist of a trained observer writing free text about behaviours that they observe, usually related to communication. This provides a qualitative assessment that may be useful for immediate feedback, but also has many limitations and inherent biases since it is necessarily

unfocussed. However, it has the advantage of the observations not being constrained by being limited to predetermined domains.

Combined methods of assessment may offer the best insight into team performance because they not only answer the questions of what happened and how, but also why they happened, which explores the motivations behind any actions observed [28]. Since the different methods offer particular advantages to different aspects of team performance, they may be combined intelligently to gain maximum benefit. For example, assessment of nontechnical skills (taskwork and teamwork) has an extensive literature using observation of behavioural markers, whereas performance distractors may be better assessed by employing events-based assessments [29]. On the other hand, communication and subtle behaviours and cognitions may be better described by field notes and self-assessment, which gain the perspectives of the observer and participant (figure 1).

FIGURE 1

6. How should we develop team assessment methods in future?

Given the burgeoning literature on assessment of team performance, it is apt to consider what ought to be the primary drivers for developing assessment tools in the future. It should be clear that assessment methods are context specific, and therefore there cannot be a single 'best assessment method', but rather there are elements of the methods that should conform to best practice [18]. A few of these best practice points will be described.

Considering the imperative for teams to learn so that they may improve, assessment should measure the processes of teamwork rather than merely the outcomes. Understanding how outcomes are achieved gives team members guidance on the impact of their specific behaviours, as well as their collective transactional behaviours. This requires assessment to move beyond mere description of the performance, towards analysis of why the team performed the way it did and took the path observed.

Unsurprisingly, the vast majority of current assessment tools base judgments on observable behaviours because they provide a quantifiable account of performance, notwithstanding the validity argument in interpreting what these behaviours mean. Also, focussing on observable and specific behaviours provides concrete grounds for feedback. However, in developing assessment tools for the future it would be advantageous to have a method that also captures unobservable cognitions. This gives further substance for debrief and promotes better understanding of underlying team processes.

The assessment must specifically address the teamwork competency being assessed, and the teamwork competencies must therefore be understood both theoretically or empirically [18]. Where possible the assessment method should distinguish between the different 'layers' of team performance, separating individual performance (taskwork) from collective behaviours (teamwork). Many healthcare curricula do not distinguish between taskwork and teamwork competencies, either

not appreciating the difference or making an assumption that they are the same thing. The effective educator or researcher needs to be more explicit if they are to approach a reliable and valid assessment method.

In a simulated environment, using specific events to trigger expected teamwork competencies is likely to provide improved reliability of assessment by reducing observer cognitive load. Similarly, a well designed tool can only perform as well as the skill and experience of the rater using it. Therefore, calibration of the assessment tool (i.e. training of the raters) is essential for optimal validity and reliability of measurement.

Finally, the reason for the assessment should be clearly understood since an assessment in isolation has no educational benefit. Education theory holds that assessment drives learning, so that in the context of a simulation-based team performance the assessment method should be tightly linked to the learning objective.

7. Summary

The assessment of team performance has grown in importance in the last few decades because of the increasing complexity of healthcare environments. There is an increasing dependence on teams to deliver optimal care, and an increased awareness of how team dysfunction contributes to error in high performance industries. Team performance depends on individual knowledge, skills and behaviours (taskwork), as well as the competence of the collective (teamwork), and assessment should distinguish between these layers. There are many assessment tools used for evaluating team performance, but they can be broadly categorised by the methodology used. The main methods employed, either singly or in combination, include the observation of behaviours, event coding, self-assessment, and field notes. Each has different advantages and disadvantages and should be chosen based on the main purpose of the assessment. No single method can be considered 'ideal' since it is context specific. In order for any assessment tool to achieve optimal validity and reliability, future development should conform to best practice guidance (Box 1).

References

1. Boreham N. A theory of collective competence: challenging the neo-liberal individualisation of performance at work. *Br J Educ* 2004;**52**(1):5-17.
2. Salas E, Frush K. Improving Patient Safety Through Teamwork and Team Training. Oxford University Press, Oxford, 2012.
3. Salas, E., DiazGranados, D., Weaver, S., King, H. (2008a). Does team training work? Principles for health care. *Academic Emergency Medicine*, 15, 1002–1009.

4. Cooke, N., Salas, E., Kiekel, P., Bell, B. (2004). Advances in measuring team cognition. In Salas, E. and Fiore, S. (Eds.), *Team Cognition: Understanding the Factors that Drive Process and Performance*. Washington, DC: American Psychological Association.
5. Institute of Medicine. (2000). *To err is human: Building a safer health system*. Available at: http://www.nap.edu/catalog.php?record_id=9728. Accessed November 2015.
6. Peter Lattman (September 27, 2007). The Origins of Justice Stewart's 'I know it when I see it'. Available at: <http://blogs.wsj.com/law/2007/09/27/the-origins-of-justice-stewarts-i-know-it-when-i-see-it/> *LawBlog at The Wall Street Journal Online*. Accessed 6 Nov 2015.
7. Heinrich HW. *Industrial Accident Prevention: A Scientific Approach*. McGraw-Hill, Columbus, 1931.
8. Reason J. Human error: models and management. *Br Med J* 2000;**320**:768–70.
9. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ* 2001;**323**:625–8.
10. Wilson T, Holt T. Complexity and clinical care. *BMJ* 2001;**323**:685–8.
11. Plsek PE, Wilson T. Complexity, leadership, and management in healthcare organisations. *BMJ* 2001;**323**:746–9.
12. Fraser SW, Greenhalgh T. Coping with complexity: educating for capability. *BMJ* 2001;**323**:799–803.
13. Paley J. The appropriation of complexity theory in health care. *J Health Serv Res Policy* 2010;**15**: 59.
14. DeVita M, Schaefer J, Lutz J, Dongilli T, Wang H. Improving medical crisis team performance. *Crit Care Med* 2004;**32**(2):S61–S65.
15. Nestel D, Walker K, Simon R, Aggarwal R, Andreatta P. Nontechnical skills: an inaccurate and unhelpful descriptor? *Simul Healthc* 2011;**6**:2–3.
16. Glavin R. Skills, training, and education. *Simul Healthc* 2011;**6**:4 –7.
17. Gaba D. Training and Nontechnical Skills: The Politics of Terminology. *Simul Healthc* 2011;**6**:8–10.
18. Rosen M, Weaver S, Lazzara E, Salas E, Wu T, Silvestri S, Schiebel N, Almeida S, King H. Tools for evaluating team performance in simulation-based training. *J Emerg Trauma Shock* 2010;**3**:353–9.
19. Baker DP, Salas E. Principles for measuring teamwork skills. *Hum Factors* 1992;**34**:469–75.
20. Weick KE, Roberts KH. Collective mind in organizations: heedful interrelating on flight decks. *Admin Sci Q* 1993;**38**:357–381.
21. Boreham N. A theory of collective competence: challenging the neo-liberal individualisation of performance at work. *Br J Edu* 2004;**52**(1):5–17.
22. Boreham N. Collective professional knowledge. *Med Educ* 2000;**34**:505–506.
23. Boreham N. *Collective competence and work process knowledge*. Paper presented to the Symposium on Work Process Knowledge in European Vocational Education and Training Research. European Conference on Educational Research, University of Crete, Greece. September 2004.
24. Lingard L. What we see and don't see when we look at 'competence': notes on a god term. *Adv Health Sci Educ Theory Pract* 2009;**14**(5):625–628)
25. Gustavsson B. Towards a transcendent epistemology of organizations. *J Organ Change Manag* 2001;**14**:352–378.
26. Brown P, Lauder H. *Human capital, social capital and collective intelligence*, in: S. Baron, J. Field and T. Schuller (Eds) *Social Capital: Critical Perspectives*, Oxford University Press, Oxford, 2000.

27. Dietz A, Pronovost P, Benson K, Mendez-Tellez P, Dwyer C, Wyskiel R, Rosen M. A systematic review of behavioural marker systems in healthcare: what do we know about their attributes, validity and application? *BMJ Qual Saf* 2014;**23**:1031-1039.
28. Rosen MA, Salas E, Wilson KA, King HB, Salisbury M, Augenstein JS, Robinson DW, Birnbach DJ. Measuring Team Performance for Simulation-based Training: Adopting best practices for healthcare. *Simul Healthc* 2008;**3**:33-41.
29. Seelandt JC, Tschan F, Keller S, Beldi G, Jenni N, Kurmann A, Candinas D, Semmer NK. Assessing distractors and teamwork during surgery: developing an event-based method for direct observation. *BMJ Qual Saf* 2014;**23**(11):918-29.
30. Wright BG, Phillips-Bute BG, Petrusa ER, Griffin KL, Hobbs GW, Taekman JM. Assessing teamwork in medical education and practice: Relating behavioral teamwork ratings and clinical performance. *Med Teach* 2009;**31**:30-8.
31. Malec JF, Torsher LC, Dunn WF, Wiegmann DA, Arnold JJ, Brown DA, Phatak V. The Mayo High Performance Teamwork Scale: Reliability and Validity for Evaluating Key Crew Resource Management Skills. *Simul Healthc* 2007;**2**:4-10.
32. Deering S, Rosen MA, Salas E, King HB. Building team and technical competency for obstetric emergencies: The Mobile Obstetric Emergency Simulator (MOES) System. *Simul Healthc* 2009;**4**:166-73.
33. Feldman M, Lazzara EH, Vanderbilt AA, DiazGranados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof* 2012;**32**:279-86.
34. Fowlkes JE, Dwyer DJ, Oser RL, Salas E. Event-based approach to training (EBAT) *Int J Aviat Psychol* 1998;**8**:209-21.
35. Rosen MA, Salas E, Wu TS, Silvestri S, Lazzara EH, Lyons R, Weaver SJ, King HB. Promoting Teamwork: An Event-based Approach to Simulation-based Teamwork Training for Emergency Medicine Residents. *Acad Emerg Med* 2008;**15**:1-9.
36. Cooper S, Cant R, Porter J, Missen K, Sparkes L, McConnell-Henry T, Endacott R. Managing patient deterioration: assessing teamwork and individual performance. *Emerg Med J* 2013;**30**(5):377-81.
37. Phitayakorn R, Minehart R, Pian-Smith MC, Hemingway MW, Milosh-Zinkus T, Oriol-Morway D, Petrusa E. Practicality of intraoperative teamwork assessments. *J Surg Res* 2014;**190**(1):22-8.
38. Schraagen JM, Schouten T, Smit M, Haas F, van der Beek D, van de Ven J, Barach P. Assessing and improving teamwork in cardiac surgery. *Qual Saf Health Care* 2010;**19**(6):e29.
39. Andrew B, Plachta S, Salud L, Pugh CM. Development and evaluation of a decision-based simulation for assessment of team skills. *Surgery* 2012;**152**(2):152-7.
40. Lie D, May W, Richter-Lagha R, Forest C, Banzali Y, Loherty K. Adapting the McMaster-Ottawa scale and developing behavioural anchors for assessing performance in an interprofessional Team Observed Structured Clinical Encounter. *Med Educ Online* 2015;**20**:26691.

41. Steinemann S, Berg B, DiTullio A, Skinner A, Terada K, Anzelon K, Ho HC. Assessing teamwork in the trauma bay: introduction of a modified "NOTECHS" scale for trauma. *Am J Surg* 2012;**203**(1):69-75.
42. McKay A, Walker S, Brett S, Vincent C, Sevdalis N. Team performance in resuscitation teams: Comparison and critique of two recently developed scoring tools. *Resuscitation* 2012;**83**(12):1478–1483.
43. Hamilton N, Freeman BD, Woodhouse J, Ridley C, Murray D, Klingensmith ME. Team behavior during trauma resuscitation: a simulation-based performance assessment. *J Grad Med Educ* 2009;**1**(2):253-9.

Table 1. Examples of assessment methods and their characteristics

Assessment category	Assessment tool	Context	Key features
Observation of Behavioural markers	Modified McMaster-Ottawa scale for rating teams [40]	<i>In vitro</i> real-time (simulation); multidisciplinary student team OSCE	6 teamworking domains: communication, collaboration, roles & responsibilities, collaborative patient-family centred approach, team conflict management, team functioning
Observation of Behavioural markers	TEAM [29]	<i>In vitro</i> real-time (simulation); nurse teams managing deteriorating patients	11 items and global assessment score over 3 domains: leadership, teamwork, task management
Observation of Behavioural markers	TEAM and OSCAR [36,42]	Retrospective <i>in vitro</i> (recorded simulation scenarios); multidisciplinary resuscitation teams	TEAM – Entire team performance OSCAR – Individuals within team
Observation of Behavioural markers	ANTS, NOTSS, SPLINTS and OTAS [37]	Retrospective video analysis of <i>in vivo</i> performance with operating theatre staff	ANTS – Anaesthetist's performance; 15 items over 4 domains: task management, situation awareness, teamwork and decision making NOTSS – Surgeon's performance; 15 items over 5 domains: situation awareness, decision making, leadership, task management, communication and teamwork SPLINTS- Nurse's performance; 9 items over 3 domains: situation awareness, task management, communication and teamwork OTAS – Entire team performance/interaction; 5 domains: communication, coordination, cooperation, leadership, team monitoring/situation awareness
Self Assessment (combined with observation of behavioural markers)	Modified CATME [39]	<i>In vitro</i> , real time (simulation). Teams of trainee surgeons performing simulated hernia repair	Self Assessment of 6 items (reduced from 87): Teamwork, Team Interaction/Communication, Team effectiveness/focus, Quality Expectations, Possession of relevant knowledge, skills and abilities & 'I would refer a family member to this team'.

Self Assessment (combined with observation of behavioural markers)	T-NOTECHS [38,41]	Trauma based: <i>In-vivo</i> real time <i>In-vitro</i> real time and retrospective video analysis	26 items over 5 domains: leadership, cooperation and resource management, communication, decision making, situation awareness and coping with stress
Event-based assessment (combined with observation of behavioural markers)	SO-DIC-OR [29]	In vivo real time in operating room	5 team-based processes: communication, leadership, problem-solving, teaching, tension
Field notes (combined with observation of behavioural markers)	Assessment of intraoperative non-routine events and team work on clinical outcomes [35]	Paediatric cardiac surgery: In vivo real time, using trained observers	Field Notes – retrospectively grouped Behavioural markers include modified versions of NOTECS, ANTS, and NOTSS (4 domains, with 13 items)
Field Notes (combined with observation of behavioural markers)	Modified Mayo High Performance Teamwork Scale [39,43]	In-vitro real time trauma simulations	Original Scale - 16 items over 4 domains: situational awareness, communication, error anticipation, error containment Modified Scale - 7 (binary) items with optional field notes

Glossary

OSCE – Observed Structured Clinical Encounter

TEAM – Team Emergency Assessment Measure

OSCAR – Observational Skill-based Clinical Assessment tool for Resuscitation

ANTS – Anesthesiologists' Non-Technical Skills

NOTSS – Non-Technical Skills for Surgeons

SPLINTS – Scrub Practitioners' List of Intraoperative Non-Technical Skills

OTAS – Observational Teamwork Assessment for Surgery

CATME – Comprehensive Assessment of Team Member Effectiveness

T-NOTECHS – Nontechnical Skills for Trauma

SO-DIC-OR – Simultaneous Observation of Distractions and Communication in the Operating Room

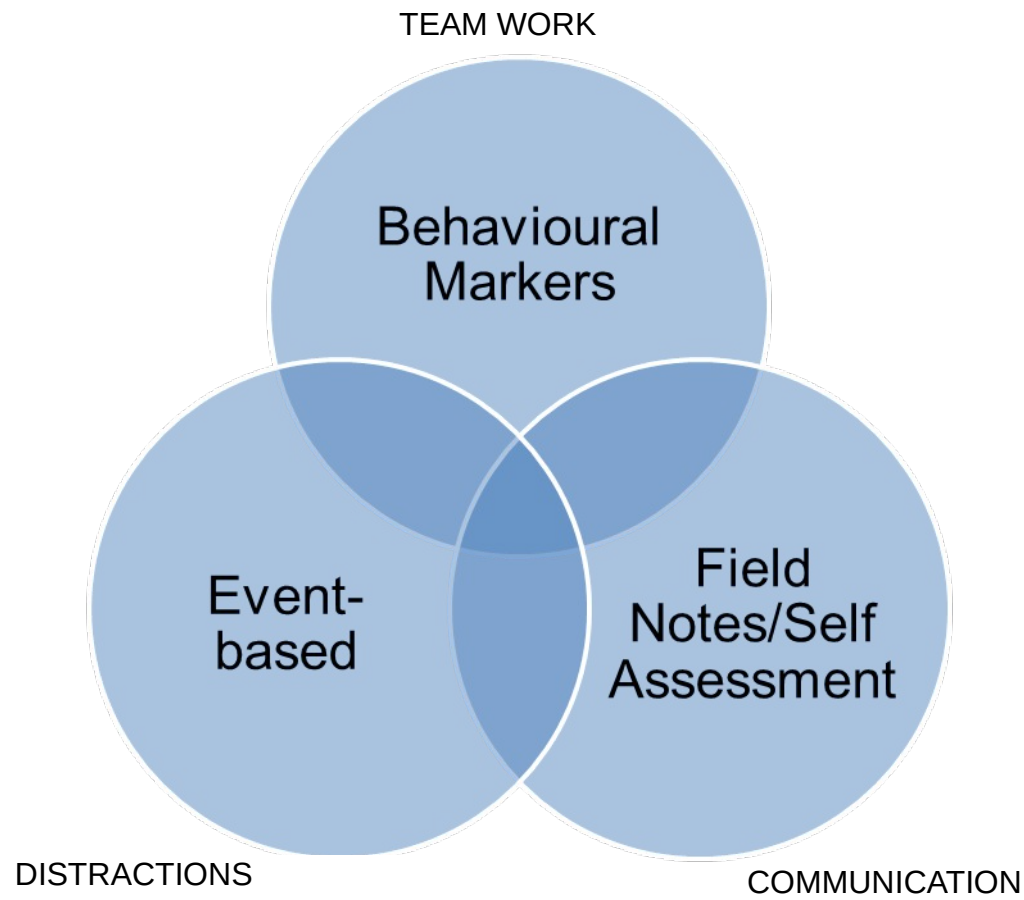


Figure 1. The circles in the Venn diagram represent the main methods for assessing team performance, with the suggested domains of teamwork they might assess. There is cross over in how the methods can be used as indicated by the intersections.

Box 1. Key considerations for developing a team assessment tool

- Measure processes of team performance rather than merely outcomes
- Move from simple description to analysis of behaviours and performance
- Try to capture tacit communications and cognitions
- Use a theory-driven or evidence-based approach to describing teamwork behaviours
- Be explicit about assessing individual or collective team behaviours
- Use specific events to trigger expected team behaviours
- Use trained and calibrated raters to optimise reliability of assessment
- Be clear about the purpose of the assessment, whether for research, learning or high-stakes decision-making

Assessing team performance – markers and methods

Highlights

- Team assessment is essential to improve performance by providing accurate feedback
- Teamwork is a collective competence which is distributed and context-dependent
- Teams can be assessed using behavioural markers, self-assessment, event-based coding, and narrative field notes
- Team assessment tools should be developed on best-practice guidelines